Actinomycete integrative and conjugative pMEA-like elements of Amycolatopsis and Saccharopolyspora decoded

Poele, Evelien M. Te; Samborskyy, Markiyan; Oliynyk, Markiyan; Leadlay, Peter F.; Bolhuis, Henk; Dijkhuizen, Lubbert

Published in:
Plasmid

DOI:
10.1016/j.plasmid.2008.01.003

Available online at www.sciencedirect.com
ScienceDirect
Plasmid 59 (2008) 202–216
PLASMID
www.elsevier.com/locate/yplas

Actinomycete integrative and conjugative pMEA-like elements of *Amycolatopsis* and *Saccharopolyspora* decoded

Evelien M. te Poele [a], Markiyan Samborskyy [b], Markiyan Oliynyk [b], Peter F. Leadlay [b], Henk Bolhuis [a,1], Lubbert Dijkhuizen [a,*]

[a] *Department of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands*
[b] *Cambridge Centre for Molecular Recognition and Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK*

Received 10 September 2007, revised 21 December 2007
Available online 3 March 2008
Communicated by Michael Chandler

Abstract

Actinomycete integrative and conjugative elements (AICEs) are present in diverse genera of the actinomycetes, the most important bacterial producers of bioactive secondary metabolites. Comparison of pMEA100 of *Amycolatopsis mediterranei*, pMEA300 of *Amycolatopsis methanolica* and pSE211 of *Saccharopolyspora erythraea*, and other AICEs, revealed a highly conserved structural organisation, consisting of four functional modules (replication, excision/integration, regulation, and conjugative transfer). Features conserved in all elements, or specific for a single element, are discussed and analysed. This study also revealed two novel putative AICEs (named pSE222 and pSE102) in the *Sac. erythraea* genome, related to the previously described pSE211 and pSE101 elements. Interestingly, pSE102 encodes a putative aminoglycoside phosphotransferase which may confer antibiotic resistance to the host. Furthermore, two of the six pSAM2-like insertions in the *Streptomyces coelicolor* genome described by Bentley et al. [Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., et al., 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141–147] could be functional AICEs. Homologues of various AICE proteins were found in other actinomycetes, in *Frankia* species and in the obligate marine genus *Salinispora* and may be part of novel AICEs as well. The data presented provide a better understanding of the origin and evolution of these elements, and their functional properties. Several AICEs are able to mobilise chromosomal markers, suggesting that they play an important role in horizontal gene transfer and spread of antibiotic resistance, but also in evolution of genome plasticity.

Keywords: Actinomycetes; Integrative and conjugative element; *Amycolatopsis*; *Saccharopolyspora*; Modular evolution

1. Introduction

Strains of the actinomycete genera *Amycolatopsis* and *Saccharopolyspora*, both belonging to the family of *Pseudonocardiaceae*, are a rich source

* Corresponding author. Fax: +31 50 3632154.
E-mail address: L.Dijkhuizen@rug.nl (L. Dijkhuizen).
[1] Present address: Department of Marine Microbiology, Centre for Estuarine and Marine Ecology, Netherlands Institute of Ecology (NIOO-KNAW), P.O. Box 140, 4400 AC, Yerseke, The Netherlands.

0147-619X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.plasmid.2008.01.003

of commercially important antibiotics, and other secondary metabolites. Well-known examples are the broad spectrum antibiotics erythromycin A produced by *Saccharopolyspora erythraea* (McGuire et al., 1952) and rifamycin produced by *Amycolatopsis mediterranei* (Sensi et al., 1959). Rifamycin is used to fight tuberculosis and leprosy. In the last decade many novel *Amycolatopsis* strains have been isolated and characterised (Tan et al., 2006; Wink et al., 2003) that may represent an unexplored source of new antibiotics and other secondary metabolites. Identification and characterisation of mobile genetic elements of *Amycolatopsis* strains is important for the construction and improvement of vectors for gene cloning and expression, and for the manipulation of *Amycolatopsis* and *Saccharopolyspora* strains producing important metabolites. Mobile genetic elements such as bacteriophages and the plasmid-like integrative elements that are discussed here are also of interest from the point of view of their own origin and evolution, and as effective modulators of host genome diversity (Frost et al., 2005).

Actinomycete integrative and conjugative elements (AICEs) can be found in strains of the genera *Amycolatopsis* and *Saccharopolyspora*. *A. mediterranei* and *Amycolatopsis methanolica* harbour pMEA100 (23.43 kb) and pMEA300 (13.3 kb), respectively, and two elements have been described for *Sac. erythraea*, pSE211 (17.2 kb) and pSE101 (10.9 kb) (Brown et al., 1988a,b). Like other AICEs, e.g. the well studied pSAM2 of *Streptomyces ambofaciens* (Smokvina et al., 1991; Boccard et al., 1989), these elements are integrated at a specific site in the host chromosome, but can also replicate autonomously (Brown et al., 1988a,b; Vrijbloed et al., 1994). Conjugal transfer of the elements, except pSE101 which lacks conjugal activity (Brown et al., 1988b), to recipient strains is accompanied by pock formation (Moretti et al., 1985; Vrijbloed et al., 1995c; Smokvina et al., 1991; Brown et al., 1988b). The integration mechanism resembles that of several temperate bacteriophages (Boccard et al., 1989), in which an integrase mediates site-specific integration into the host chromosome in a conserved tRNA gene. Recombination occurs between an identical short sequence in the attachment site that is present on the element (*attP*) and on the chromosome (*attB*), where it overlaps the 3′ end of a specific tRNA. A general feature of most AICEs is a conserved organisation of the genes *xis*, *int* and the *attP* site (Brown et al., 1994) directly downstream of the *rep* gene (te Poele et al., 2006).

Recently, we have shown that pMEA100, pMEA300, and pSE211 form a novel group of AICEs, based on the unique characteristics of their replication initiator protein RepAM, that does not display similarity to any previously known replication proteins (te Poele et al., 2006). Binding studies with purified pMEA300 RepAM protein revealed that it is able to bind to multiple identical 8 bp repeats within its own *repAM* coding sequence. The repeat sequences within this putative origin of replication (*ori*) can form a stable secondary, hairpin structure. Similar structures with multiple identical 8 bp inverted repeats were found at the 3′ end of the putative replication initiator genes of pMEA100 and pSE211.

Several pMEA-like sequences, such as the novel *repAM* genes, were found to be widely present in strains of the genus *Amycolatopsis* (te Poele et al., 2007). Phylogenetic analysis of deduced pMEA-related RepAM and TraJ protein sequences revealed clustering with the protein sequences of pMEA300 or pMEA100. Two distinct populations of pMEA-like elements could be distinguished, one found in isolates from Europe and the other in isolates from Australia and Asia. Linkage between the distribution of *repAM* and *traJ* and the 16S rRNA gene revealed co-evolution of the elements with their hosts and suggested that they evolved in an integrated form rather than by horizontal gene transfer of the freely replicating form.

Previously, we have analysed the complete nucleotide sequence of pMEA300 of *A. methanolica* (GenBank Accession No. L36679) (Vrijbloed et al., 1994, 1995c,a,b). Recently, also the complete genome sequence of *Sac. erythraea* NRRL23338 has been determined (Oliynyk et al., 2007), which included the AICEs pSE211 and pSE101. Here we report the complete nucleotide sequence of pMEA100 of *A. mediterranei* and present a structural and evolutionary comparison to the other pMEA-like elements, to other AICEs, and to two newly discovered elements in the genome sequence of *Sac. erythraea*. This comparative analysis provides a better understanding of the origin and evolution of these elements, and their functional properties.

2. Material and methods

2.1. pMEA100 cloning

Plasmid pMEA101 (pMEA100 cloned into the *SphI* site of pBR322) (Moretti et al., 1985) was used for nucleotide sequencing. Based on the available restriction map (Moretti et al., 1985) pMEA101 was restricted with *Bam*HI resulting in five fragments of 10.4, 7.1, 6.1, 3.0, and 0.8 kb. The 7.1 kb fragment contained the cloning vector pBR322 and was circularised by self-ligation. The other four fragments were ligated into the *Bam*HI site of pBluescript II KS-(Stratagene).

2.2. DNA sequencing and analysis

DNA sequencing was done by primer walking and performed by GATC GmbH (Konstanz, Germany), Agowa GmbH (Berlin, Germany) and at our in house sequencing facility. Sequencing difficulties in DNA regions with strong secondary structures, or with a high percentage of G and C nucleotides, were overcome by doubling the concentrations of the components of the sequencing reaction mixture, except those of the primers and DNA template. Sequence reads were assembled and analysed using SeqMan (Lasergene 7.0) and Chromas 2.23. Putative open reading frames (ORFs) were identified with Clonemanager 6.0 software. Searches for homologues of the putative ORFs were performed with BLAST (Altschul et al., 1997). DNA and amino acid sequences were aligned using Clonemanager 6.0 software and ClustalW (http://www.ebi.ac.uk/clustalw/). The TMHMM software (version 2.0) (http://www.cbs.dtu.dk/services/TMHMM) was used to predict transmembrane helices in putative proteins. Secondary structures of DNA were predicted by mfold (Zuker, 2003). The complete nucleotide sequence of pMEA100 was deposited in the GenBank database under Accession number EU149765. Phylogenetic trees of pMEA100-encoded putative proteins and homologues were constructed using the neighbour-joining algorithm of Mega version 4.0 (http://www.megasoftware.net/mega4/index.html) (Tamura et al., 2007).

3. Results and discussion

After assembling all DNA sequences from pMEA100, we obtained 1 final contig of 23290 bp. Sequencing primers were used to bridge the *Bam*HI sites used in the subcloning. This revealed an additional *Bam*HI (12832 bp) site at only 24 bp of the previously identified *Bam*HI (12808 bp) site, which was not detectable by restriction mapping. pMEA100 has an average G+C content of 68.5%, which is similar to that of the other pMEA-elements, i.e. 69.3% for pMEA300 and 68.9% for pSE211. The first 6500 bp contains regions with a G+C content that is significantly below average (Fig. 1A) which may indicate that these regions have been acquired by horizontal gene transfer. ORF prediction revealed 27 ORFs in the pMEA100 sequence (Table 1 and Fig. 1B). In general, closest homologues of the pMEA100-encoded ORF products were found within the actinomycetes, except for seven proteins with no significant homologues in databases and one protein (Amp111) showing highest similarity to a *Ralstonia eutropha*-encoded protein.

Analysis of the *Sac. erythraea* genome sequence revealed the presence of several pMEA100 homologues, namely the previously identified pSE211 and pSE101 elements, plus two other putative AICEs (Fig. 1B). The first of these new elements (~20.4 kb; from bp 1245945 to bp 1266370) was designated pSE222, encoding proteins displaying clear amino acid sequence similarity with proteins of the pMEA-like elements and other AICEs. The second integrated element (~11.7 kb, from bp 273573 to ~bp 285239) is remarkably similar to pSE101, and therefore was designated pSE102. The structural and functional properties of these novel elements are discussed here as well.

The genome of *Streptomyces coelicolor* A3(2) contains six pSAM2-like insertions, three of which are integrated into a tRNA gene (Bentley et al., 2002; this study). Analysis of these three insertions showed that two of them are putative functional AICEs. These AICEs, designated AICE*Sco*3250 (14.3 kb) and AICE*Sco*5349 (21.2 kb) (Fig. 1D), are included in the comparison to the pMEA-elements. The other three insertions appear to be AICE remnants and are not discussed here.

One of the most striking features when comparing the genetic maps of these *Amycolatopsis* and *Saccharopolyspora* AICEs is their highly conserved structural organisation (Fig. 1B). Four functional modules could be distinguished on each of the elements, i.e. replication, excision/integration, regulation, and conjugative transfer. The conserved structural organisation of the pMEA-related elements can also be found on three other sequenced AICEs, SLP1 of *S. coelicolor* A3(2), pSAM2 of *S. ambofaciens* and pMR2 of *M. rosaria* (Fig. 1C) and on the two AICEs of *S. coelicolor* (Fig. 1D).

Analysis of the evolutionary relationship of AICE-encoded proteins shows that several pMEA-encoded proteins are highly related, such as the replication proteins of pMEA300, pSE211 and

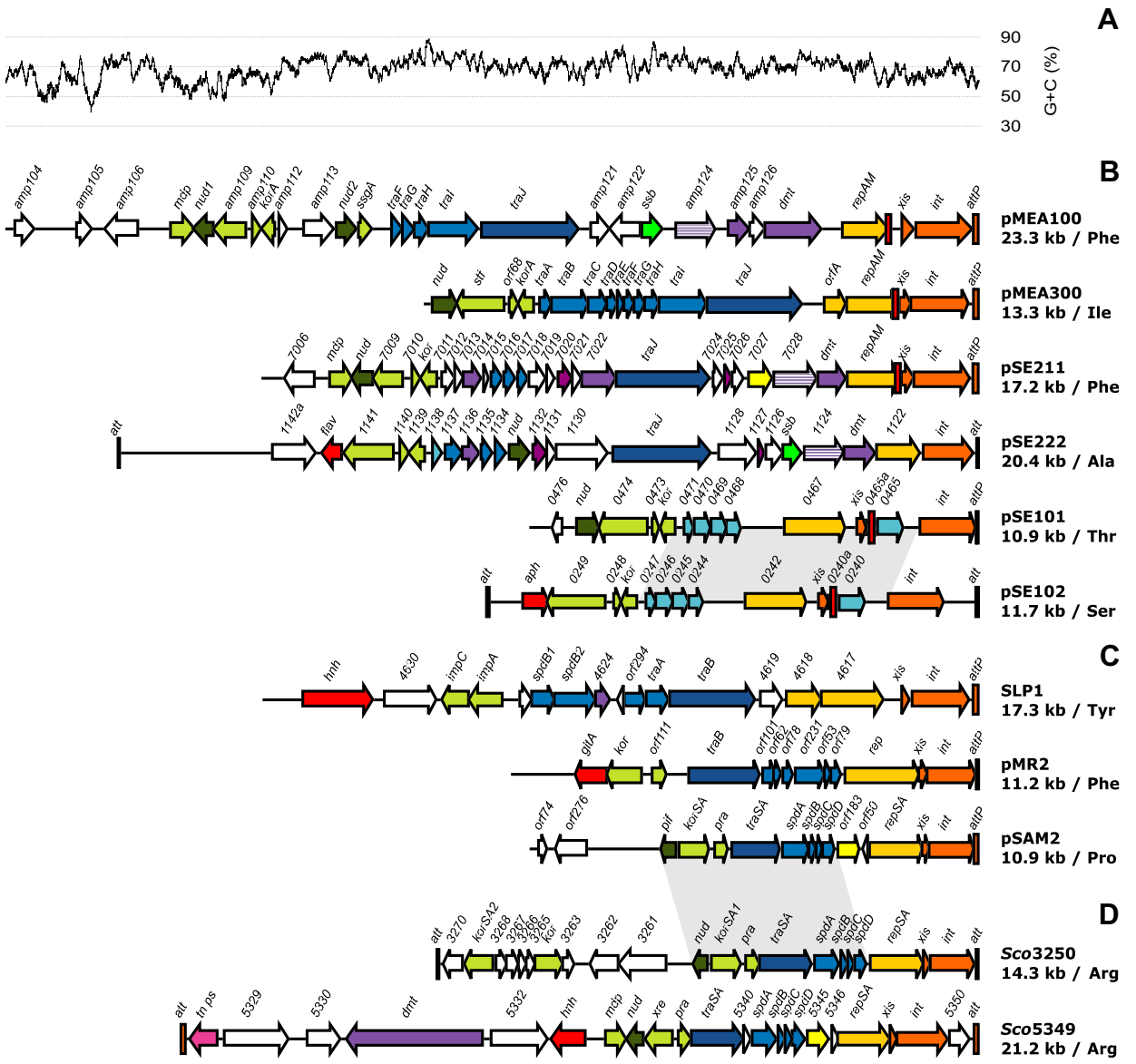

Fig. 1. (A) G+C content (%) of pMEA100 of *A. mediterranei*; structural organisation of (B) the pMEA-elements pMEA100 of *A. mediterranei*, pMEA300 of *A. methanolica*, and pSE211 of *Sac. erythraea* NRRL23338, and the other AICEs of *Sac. erythraea* pSE222, pSE101, and pSE102 and (C) the AICEs SLP1 of *Sac. coelicolor* A3(2), pMR2 of *M. rosaria*, and pSAM2 of *S. ambofaciens*. (D) newly found AICEs AICE*Sco*3250 and AICE*Sco*5349 of *S. coelicolor* A3(2). The size of the elements and the tRNA gene in which the elements are inserted are indicated below the element name at the right. Colour coding: orange, genes and sites involved in excision/integration; dark yellow, genes most likely involved in replication and its control; red bar, pMEA-specific hairpin structure; light blue, putative conjugation genes; dark blue, putative main transfer genes; lime, putative regulatory genes; dark green, Nudix hydrolase genes; white, *orfs* with unknown functions; pink, transposon pseudogene (*tn ps*); lavender with white diagonal lines, genes encoding DNA primase/polymerase (Prim-pol) proteins; red, genes with annotated function: *gltA*, glycosyltransferase; *hnh*, HNH-endonuclease signature; *aph*, aminoglycoside phosphotransferase; *flav*, flavoprotein. The following colours correspond to genes shared by two or more elements, lavender, pMEA100, pSE211, SLP1, AICE*Sco*5349; bright yellow (GGDEF-domain), pSE211, pSAM2, AICE*Sco*5349; aqua, pSE101, pSE102, pSE222; bright green (single-stranded binding protein), pMEA100 and pSE222; plum, pSE211, pSE222. Grey band between two elements indicates highly similar DNA regions.

pMEA100 and the excisionases and integrases of pMEA100, pSE211, pSE101 and pSE102 (Fig. 2). However, many of the other proteins show weak or no significant sequence similarity to each other. This mosaic-like, modular structure is typical for mobile genetic elements as they evolve by exchange

Table 1
ORFs and genetic elements identified on pMEA100 from *A. mediterranei*, and the closest homologue of each pMEA100-encoded ORF product found with BLAST searches

ORF No.	Position (bp)	Gene	Length (aa)	Closest BLAST match	percentage identity	*E*-value	Accession No.
AMP101	1–1113	*repAM*	370	429 aa hypothetical protein Mmcs_4162 of *Mycobacterium* sp. MCS	53% (172/319)	2,00E-79	YP_641323
	1121–1299	*hairpin*					
AMP102	1434–1736	*xis*	100	98 aa putative excisionase of pSE211 of *Sac. erythraea* NRRL 23338	65% (39/60)	2,00E-10	P22876
AMP103	1752–3098	*int*	448	bacteriophage-like integrase (437 aa) of pSE211 of *Sac. erythraea* NRRL 23338	63% (275/435)	1,00E-141	YP_001109118
	3134–3283	*attP*	n.a.				
AMP104	3394–3855	*amp104*	153	Not found			
AMP105	4891–5238	*amp105*	115	Not found			
AMP106	6346–5552C	*amp106*	264	Not found			
AMP107	7116–7688	*mdp*	190	183 aa metal dependent phosphohydrolase of pSE211 of *Sac. erythraea* NRRL 23338	65% (116/177)	4,00E-56	YP_001109093
AMP108	8156–7683C	*nud1*	158	158 aa NUDIX hydrolase of pSE211 of *Sac. erythraea* NRRL 23338	75% (118/158)[a]		
AMP109	8911–8165C	*amp109*	249	247 aa XRE family transcriptional regulator of pSE211 of *Sac. erythraea* NRRL 23338	58% (144/249)[a]		
AMP110	9084–9284	*amp110*	67	68 aa ORF68 of pMEA300 of *A. methanolica*	37% (25/67)[a]		
AMP111	9631–9284C	*korA*	115	GntR motif. 235 aa regulatory protein GntR, HTH-GntR, C-terminal of *R. eutropha* JMP134	38% (36/94)	5,00E-06	YP_295475
AMP112	9752–10009	*amp112*	85	Not found			
AMP113	10318–11040	*amp113*	240	Not found			
AMP114	11081–11602	*nud2*	174	244 aa putative DNA hydrolase protein pRL1.17 of pRL1 of *Streptomyces* sp. 44030	51% (70/137)	8,00E-25	ABC67346
AMP115	11623–11976	*ssgA*	117	133 aa putative SsgA-family transcriptional regulator of *F. alni* ACN14a	46% (51/110)	4,00E-13	YP_715692
AMP116	12445–12675	*traF*	76	79 aa TraF of pMEA300 of *A. methanolica*	49% (39/79)[a]	—	2124370F
AMP117	12675–12989	*traG*	104	84 aa TraG of pMEA300 of *A. methanolica*	53% (34/64)	8,00E-05	2124370G
AMP118	12982–13314	*traH*	110	115 aa TraH of pMEA300 of *A. methanolica*	35% (38/108)	1.1	2124370H
AMP119	13311–14492	*traI*	393	352 aa hypothetical protein pRL1.11 of pRL1 of *Streptomyces* sp. 44030	43% (110/251)	6,00E-30	ABC67340
AMP120	14599–16911	*traJ*	770	FtsK/SpoIIIE domain. 616 aa putative plasmid transfer protein of *S. tenjimariensis*	56% (340/598)	7,00E-159	CAH60136
AMP121	17208–17654	*amp121*	148	Not found			
AMP122	18387–17644C	*amp122*	247	Not found			
AMP123	18442–18924	*ssb*	160	141 aa single-strand DNA-binding protein of SACE_1125 of pSE222 of *Sac. erythraea* NRRL 23338	50% (82/161)	2,00E-30	YP_001103381
AMP124	19233–20187	*amp124*	318	314 aa hypothetical protein SACE_1124 of pSE222 of *Sac. erythraea* NRRL 23338	37% (114/303)	8,00E-31	YP_001103380
AMP125	20523–21023	*amp125*	166	141 aa hypothetical protein SACE_1136 of pSE222 of *Sac. erythraea* NRRL 23338	41% (47/113)	7,00E-10	YP_001103392
AMP126	21020–21373	*amp126*	117	246 aa hypothetical protein SCP1.122 of SCP1 of *S. coelicolor* A3(2)	50% (52/104)	6,00E-20	NP_639695
AMP127	21373–22720	*dmt*	541	Cyt_C5_DNA_methylase domain. 541 aa putative DNA methyltransferase of *N. farcinica* IFM 10152	51% (283/548)	1,00E-137	YP122127

[a] Calculated using ClustalW (Blosum) percentage of matching amino acids.

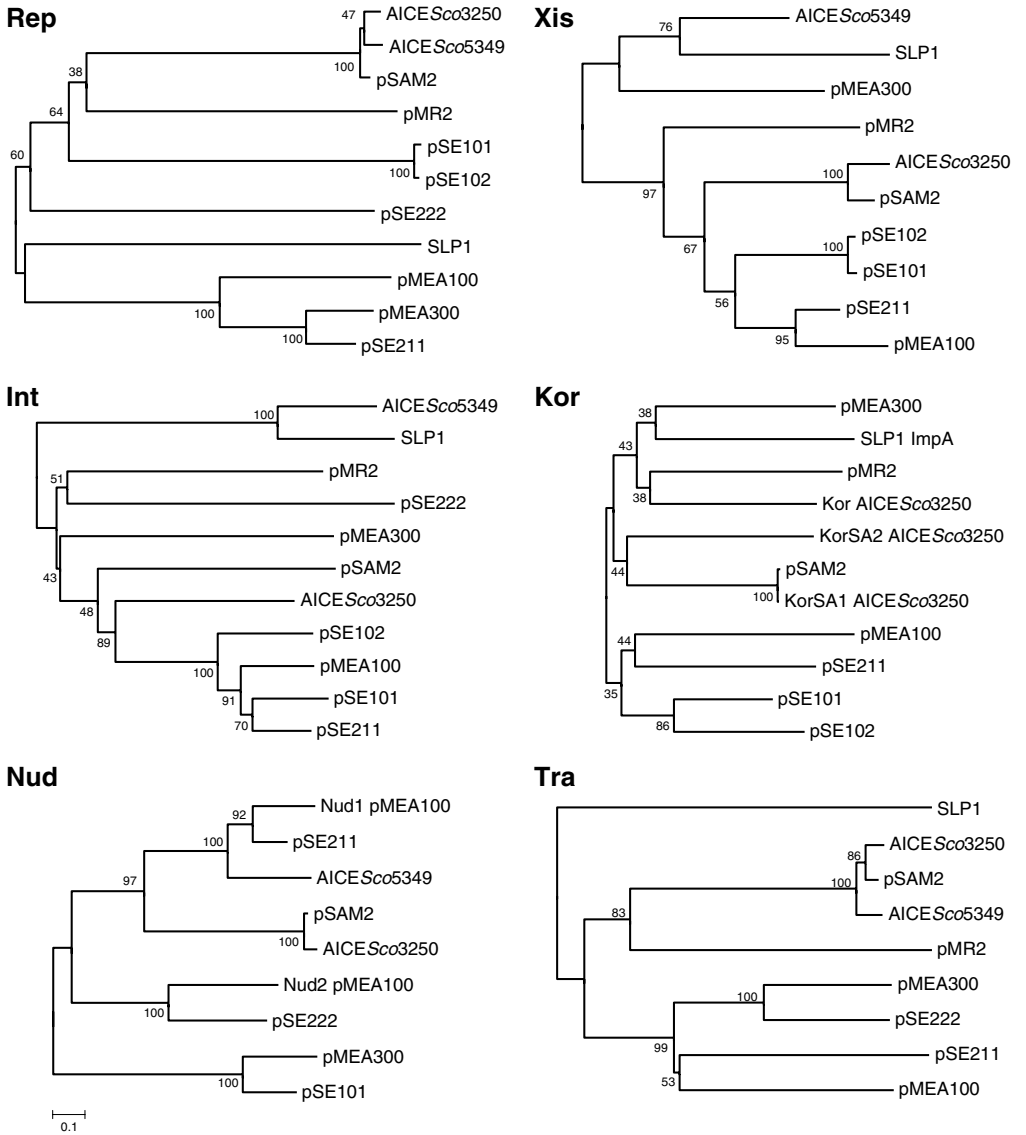

Fig. 2. Evolutionary relationship between several AICE–encoded proteins; Rep, replication proteins; Int, integrases; Xis, excisionases; Nud, Nudix hydrolases; Kor, Kor proteins; and Tra, main transfer proteins. The scale bar represents 0.1 substitutions per site.

of functional modules from related and different types of mobile genetic elements (Toussaint and Merlin, 2002).

The observed mosaicism of AICEs suggests that, either the elements are functionally related, but have a different origin (convergent evolution), or that the elements had a common ancestor from which they diverged early in the evolution of these elements. The observation that several elements share highly related proteins suggests that exchange of modules or genes has occurred between the elements.

3.1. Replication

The *repAM* gene of pMEA100 encodes the putative replication initiator protein and is related to the RepAM proteins of pMEA300 and pSE211 (te Poele et al., 2006). RepAM furthermore shows 53% identity to a hypothetical protein of *Mycobacterium* sp. strain MCS (Table 1) and 43% (158/364) to a possible prophage ΦRv2 protein (NP_217171) of *M. tuberculosis*. The putative *ori* of pMEA100 is located near the 3′ end of the *repAM* gene and contains a strong 178 bp hairpin structure. How-

ever, unlike the hairpins on pMEA300 and pSE211, the pMEA100 structure is located outside *repAM*, namely within the intergenic region between *repAM* and the putative excisionase (*xis*). Repeated sequencing directly on genomic DNA isolated from the *A. mediterranei* wild type consistently gave a TGA stop codon at 1113 bp of *repAM*, showing that *repAM* of pMEA100 is shorter than its homologues on pMEA300 and pSE211. The shorter *repAM* gene of pMEA100 most likely encodes a truncated but functional protein. However, the possibility that it is a pseudogene that no longer encodes a functional protein, can not be excluded.

The pSE101, pSE102 and pSE222 elements lack a *repAM*-related gene and apparently encode other replication genes instead. In that sense, they are not genuine pMEA-elements, but are clearly pMEA-related, with a similar overall organisation, and several other common genes and structural features. The protein products of the pSE101 and pSE102-encoded putative replication genes (SACE_0242 and SACE_0467) are highly related to each other (92% (452/488) identity) and show 34% identity (120/350) to a predicted ATPase protein (YP_001236766) involved in replication control of *Bradyrhizobium* sp. BTAi1. The pSE101- and pSE102-encoded proteins could therefore be related to helicases, which have ATPase activity as well as unwinding activity. Interestingly, pSE101 and pSE102 both contain the pMEA-specific hairpin that is highly similar (80 and 86%, respectively) to the 174 bp hairpin structure of pSE211. The hairpin of pSE101 (167 bp; 63 kcal/mol) is 98% identical to the one of pSE102 (122 bp; $\Delta G = 46$ kcal/mol), except for a 45 bp stretch (bp 45 to bp 95 from the pSE101 hairpin) that is missing from the pSE102 hairpin. The hairpins are located directly downstream of the putative *xis* genes and also contain the multiple 8 bp repeats and putative nicking sites as described for pMEA300. The pMEA300-encoded RepAM is able to bind to the multiple identical 8 bp repeats in this putative *ori* (te Poele et al., 2006). However, since the replication proteins of pSE101 and pSE102 are clearly distinct from the pMEA-encoded RepAM proteins, it remains unclear whether they also use the conserved hairpin structure as *ori* or whether they use an alternative *ori*. Alternatively, the hairpin structures of pSE101 and pSE102 may be recognised by a host-encoded protein or a protein encoded by another element, like pSE211.

The pSE222 element most likely encodes a replicative helicase (SACE_1122; 349 aa) that belongs to the family of RecA-like NTPases. This family includes the NTP-binding domain of DnaB and related helicases. Its amino acid sequence is highly related (64% identity (194/299)) to a putative replicative helicase RepA (YP_001156923) of *Salinispora tropica* CNB-440. Furthermore, the RecA-like NTPase domain of SACE_1122 shows 30% (60/198) identity to the C-terminal helicase domain of Gp57 (YP_655334) of Mycobacteriophage Pipefish. This 871 aa protein, which is more than two times larger than SACE_1122, has an additional N-terminal Prim-pol domain. This domain is absent from SACE_1122, but is encoded by another gene (SACE_1124) upstream of the helicase (see below). Proteins with such a domain are known to harbour both DNA-primase and polymerase activity (Lipps et al., 2003) and are often fused to helicases or are encoded by genes in the proximity of genes encoding helicases (Iyer et al., 2005). It has been suggested that the Prim-pol proteins with the associated helicases could form a replication initiation complex (Lipps, 2004). The pSE222-encoded putative Prim-pol protein and putative helicase, which are separated by a *dmt* gene, could therefore form the replication complex of pSE222.

Homologues of the pSE222-encoded putative Prim-pol protein can be found on pMEA100 (Amp124; 318 aa) and pSE211 (SACE_7028; 337 aa) showing 37% (114/303) and 30% (87/285) identity, respectively, to Prim-pol of pSE222. The pMEA-encoded Prim-pol proteins also show ~30% similarity to the N-terminal Prim-Pol domain of a 955 aa long phage/plasmid primase P4-like protein of *Sphingomonas wittichii* RW1. This protein has an additional C-terminal helicase domain, which is missing from the much smaller pMEA-encoded Prim-pols. As opposed to pSE222, pMEA100 and pSE211 do not encode proteins with an obvious helicase activity. Whether their Prim-pol proteins are involved in replication is not clear.

3.2. Site-specific integration and excision

The second module is involved in site-specific integration of the elements and contains the excisionase, integrase and *attP* site. The putative excisionase *xis* of pMEA100 (*amp102*) encodes a 100 aa protein resembling the excisionases of pSE211, pSE101 and pSE102 (65% (39/60), 48% (25/52) and 48% (25/52) identity, respectively). An excision-

ase homologue appears to be absent on pSE222. Conceivably, its integrase can carry out both integration and excision, as shown for the integrase of SLP1 (Brasch and Cohen, 1993). The integrases of SLP1 and pSE222 only show 23% (84/352) identity to each other. The putative integrase Int of pMEA100 is encoded by *amp103* and shows extensive similarity to the integrases of pSE211, pSE101 and pSE102 (63% (275/435), 58% (265/451) and 45% (214/467) identity, respectively). The pSE222 integrase is highly similar (72% (291/400) identity) to a phage integrase (Sare_1922; YP_001536798) of *Salinispora arenicola* and has 27% (97/350) identity to the integrase (YP_220445) of pMR2, the AICE of *Micromonospora rosaria*. Interestingly, the pSE222-encoded integrase is also highly related (69% (265/381) identity) to an integrase (SCO6806) of *S. coelicolor* encoded on a large chromosomal region of 148 genes, that was probably acquired by horizontal gene transfer and includes genes for heavy metal resistance and secondary metabolite production (Bentley et al., 2002).

The pMEA-encoded integrases and the integrases of the other AICEs are site-specific tyrosine recombinases related to bacteriophage integrases (Boccard et al., 1989). The attachment site *attP* (150 bp) of pMEA100 can be found 36 nucleotides downstream of *int* and pMEA100 is integrated, like pSE211 (Brown et al., 1990), in an *attB* site that corresponds to a Phe-tRNA gene (Madon et al., 1987). pMEA300 integrates into an Ile-tRNA gene (Vrijbloed et al., 1994) and pSE101 is integrated in a Thr-tRNA gene (Brown et al., 1994). The putative attachment sequence (*att*) flanking both sites of the integrated pSE222 element is 45 bp long. One sequence is located within the 3′ end of the Ala-tRNA directly downstream of the pSE222 integrase gene. Similarly, the putative *att* of pSE102 is 55 bp long, and is located within a Ser-tRNA directly downstream of *int*.

3.3. Regulation

The pMEA100 *amp107* to *amp111*, *amp114* and *amp115* encode homologues of regulatory proteins. The *mdp* (*amp107*)-encoded protein has a hydrolase domain (HD domain), which is found in a superfamily of enzymes with phosphohydrolase activity. The presence of highly conserved histidine and aspartate residues in the HD superfamily suggests that divalent cations are essential for activity. These metal-dependent phosphohydrolase enzymes may be involved in nucleic acid metabolism and signal transduction (Aravind and Koonin, 1998). pSE211 also encodes a putative metal-dependent phosphohydrolase that is highly related to the pMEA100 homologue, Amp107 (Table 1).

Interestingly, two *orf*s (*amp108* and *amp114*) on pMEA100 encode putative Nudix hydrolase proteins, designated Nud1 and Nud2, respectively. Nudix proteins hydrolyse the pyrophosphate bond in a Nucleoside diphosphate linked to some other moiety, X. The *nud1* gene, located directly downstream and transcribed in the opposite direction from *mdp*, encodes a protein of 158 aa and is highly similar (75% identity) to the putative Nudix hydrolase of pSE211 (Table 1). Nud2 consists of 174 aa and is 52% (68/130) identical to a putative Nudix hydrolase encoded by pSE222. pMEA300 and pSE101 also encode a putative Nudix hydrolase, i.e. ORF192 and Nud101, respectively. They show high similarity to each other (68% (130/192) identity) but no significant similarity to the Nudix hydrolases of pMEA100 and pSE211. The pMEA100 Nudix proteins are also not related to each other, suggesting that they were acquired from different ancestors, rather than being the result of a duplication event. A well-known Nudix protein is MutT of *E. coli*, which degrades potentially mutagenic nucleotides. Other Nudix proteins control the levels of metabolic intermediates and signalling compounds (McLennan, 2006). The *pif* gene of pSAM2 encodes a Nudix hydrolase (Pif) that confers conjugal immunity. Via Pif, redundant exchange of pSAM2 between donor cells is prevented and the element remains integrated in the chromosome (Possoz et al., 2003). The pSAM2 protein lacks any known DNA-binding or transmembrane domains, but requires the Nudix domain for activity. It is therefore suggested that Pif most likely prevents recognition between donors by modifying a host component in the donor strain (Possoz et al., 2003). Pif is related to the Nud1 of pMEA100 (28% (44/158) identity) and Nud of pSE211 (29% (46/158) identity), but not to Nud2 of pMEA100 and the Nudix proteins of pMEA300, pSE101 and pSE222. A Nudix hydrolase gene is absent on pSE102.

amp109 is located directly downstream of *nud1* and encodes a 249 aa protein with an N-terminal helix-turn-helix (HTH) DNA-binding XRE-motif. Proteins with such a motif belong to the xenobiotic response element family of transcriptional regulators. A highly similar protein of 247 aa encoded

by SACE_7009 was found on pSE211, showing 58% identity to the pMEA100-encoded regulator, and is also located directly downstream of the Nudix hydrolase gene (*nud211*).

Both pMEA300 and pSE101 also encode proteins (Stf and SACE_0474) with 28% (104/367) identity to a XRE-family transcriptional regulator (YP_949921) of *Arthrobacter aurescens* and their genes are also located directly downstream of the nudix hydrolase gene. However, these proteins lack the N-terminal HTH-XRE motif and do not resemble the pMEA100- and pSE211-encoded proteins. Deletion of the pMEA300-encoded *stf* gene was shown to result in reduced transformation frequencies with pMEA300 DNA in *A. methanolica* (Vrijbloed et al., 1995b). SACE_0474 of pSE101 encodes a protein of 392 aa and is 70% (227/324) identical to the 380 aa Stf protein of pMEA300. Furthermore, both proteins contain a partial C-terminal DUF921 domain, which is absent on the pMEA100 and pSE211 XRE-proteins. DUF921 is found in several putative regulatory proteins in *Streptomyces* and other actinomycetes, one of which also has a putative N-terminal HTH motif and is thought to be involved in sporulation regulation, i.e. *orf1590* (P19471) of *Streptomyces griseus* (Babcock and Kendrick, 1990). pSE102 encodes a protein (SACE_0249; 471 aa) with 27% identity to the sporulation associated protein of *S. griseus* as well, but shows no significant similarity to the proteins of pMEA300 and pSE101. The pSE222 element encodes a protein (SACE_1141) with an N-terminal HTH-XRE domain, which is not related to the pMEA-encoded XRE-proteins or to the putative sporulation regulators, but has homologues in *Sal. tropica* (YP_001160170; 34% (141/407) identity) and *Sal. arenicola* (ABV99397; 34% (143/409) identity).

A highly similar *mdp*, *nud*, *xre* cluster is present on AICE*Sco*5349, showing 57% and 62% DNA identity to the corresponding gene clusters of pMEA100 and pSE211. At the amino acid level, the putative Mdp (SCO5335), Nudix hydrolase (SCO5336) and HTH-XRE-regulator (SCO5337) are 52 and 54%, 62 and 70%, 46% and 46% identical to the corresponding proteins of pMEA100 and pSE211, respectively.

amp110 encodes a small hypothetical protein of 67 aa that is conserved amongst the pMEA-related elements, i.e. ORF68 of pMEA300 (68 aa; 37% identity), SACE_7010 of pSE211 (66 aa; 34% identity), SACE_0473 of pSE101 (59 aa; 13% identity), SACE_0248 of pSE102 (59 aa; 22% identity) and SACE_1140 of pSE222 (66 aa; 27% identity). Interestingly, these small *orfs* are located at similar positions on the elements, i.e. directly upstream and convergently transcribed from *korA* (see below) on the pMEA-elements and from a second XRE-family regulator (SACE_1139) on pSE222. TMHMM analysis predicts a signal peptide for all six proteins, which indicates that they may be secreted. No other Amp110 homologues were found in the databases, suggesting that these proteins are specific for the pMEA-related elements.

Amp111 (KorA, 115 aa) contains a DNA-binding GntR domain and has homologues on all pMEA-elements. The pMEA300 homologue (118 aa) is 29% identical to KorA of pMEA100 and is involved in regulation of replication and conjugal transfer (Vrijbloed et al., 1995c). KorA proteins are also present on other AICEs, KorR (274 aa) of pMR2, ImpA (269 aa) of SLP1, and KorSA (259 aa) of pSAM2. Interestingly, one of the two newly found AICEs of *S. coelicolor*, AICE*Sco*5349 does not encode a Kor protein, whereas AICE*Sco*3250 encodes three Kor-like proteins (Fig. 1). The proteins are only 24-27% identical to each other, but one of the proteins (SCO3259) is strikingly similar (98% identity) to KorSA of pSAM2. The gene encoding this KorSA homologue is located on a ~4.5 kb DNA region that is ~80% similar to a DNA region of pSAM2.

The pSAM2-encoded KorSA negatively regulates transfer by repressing *pra*, which encodes the activator of replication and transfer (Sezonov et al., 2000). KorSA contains an additional C-terminal domain (UTRA) that is able to bind small effector molecules. ImpA of SLP1 is 28% identical to KorSA and also has a UTRA domain. The *impA* gene together with the translationally coupled *impC*, constitute the *imp* (inhibition of plasmid maintenance) locus of SLP1 and also prevents replication and transfer of the integrated SLP1 (Shiffman and Cohen, 1993). GntR domains are often located at the N-terminal part of transcriptional regulators with a variable effector-binding/oligomerisation domain at the C-terminus that modulates the activity of the DNA-binding protein (Rigali et al., 2002). The pMEA-related Kor proteins, however, are approximately half the size of other AICE Kor proteins and do not have an additional known conserved domain besides the DNA-binding domain. KorA of pMEA100 shows 38% identity to the N-terminal GntR-containing part of a regulatory

protein of *Ralstonia eutropha* JMP134. This protein is approximately 2 times larger (235 aa) than KorA of pMEA100 and contains an additional C-terminal FadR domain, which is often found in regulators of fatty acid metabolism (Rigali et al., 2002). pSE222 lacks a GntR transcriptional regulator, but encodes a second XRE-family regulator (SACE_1139) instead.

Several AICE-encoded *kor* genes are also part of a *kil-kor* system. The expression of certain (*kil*) genes is lethal in the absence of the *kor* (*kil* override) gene that controls the expression of the Kil-phenotype. The *kil-kor* systems of the AICEs, like the *kil-kor* systems found on conjugative and pock forming *Streptomyces* plasmids, are associated with conjugation, in which Kor transcriptionally represses transfer genes responsible for the Kil-phenotype. The proposed function is to retard growth of recipient cells, resulting in pock formation, until the copy number of the AICE is sufficiently high (Hagege et al., 1993).

amp115 encodes a protein (117 aa) with similarity to SsgA-like protein (SALP) transcriptional regulators showing 46% (51/110) identity to a SsgA-family homologue of *Frankia alni* ACN14a. This group of regulators is present in sporulating actinomycetes. In *S. coelicolor* A3(2) SsgA plays an important role in sporulation and cell division (van Wezel et al., 2000). Its function on pMEA100 remains unclear and no homologues were found on the other pMEA-elements or AICEs, but only in actinomycete genomes.

3.4. Conjugative transfer

The *amp116-amp120* (*traFGHIJ*) gene cluster of pMEA100 encodes proteins involved in conjugal transfer. The genes of homologues of these proteins are found on pMEA300 on which they are part of a larger conjugation cluster (*TraA* to *TraJ*). TraA and TraB of pMEA300 are responsible for the pock phenotype and TraG to TraJ are required for efficient transfer of pMEA300 to other cells (Vrijbloed et al., 1995c). Since TraA and TraB-homologues are missing from pMEA100, other proteins encoded by pMEA100 must be responsible for the observed pock phenotype in *A. mediterranei* (Moretti et al., 1985). Interestingly, pSE211 and pSE222 have, respectively, 3 and 2 copies of the *traB* gene, but lack the other pMEA300-like transfer genes, except *traJ*, the main transfer gene (see below). The pMEA100-encoded TraF, TraG, and TraH proteins have 49, 53 and 35% identity to homologues from pMEA300 (Table 1). Only the C-terminal regions of TraI of pMEA100 and pMEA300 show similarity (43% (19/44) identity) to each other, but they have several other features in common. The *traI* genes of pMEA100 and pMEA300 are located directly upstream of *traJ* and the *traI* start codon overlaps the stop codon of *traH*, the gene directly upstream of *traI*. Moreover, both TraI proteins are putative transmembrane proteins with three (pMEA100) and four (pMEA300) predicted transmembrane helices (TMHMM). TraJ of pMEA100 shows high similarity (56% (340/598) identity) to a putative plasmid transfer protein of *Streptomyces tenjimariensis* and 34% (263/752) identity to the main conjugation protein TraJ of pMEA300. Homologues of TraJ are also present on pSE211 and pSE222. The TraJ proteins all have two transmembrane helices in the N-terminal part of the protein, except for the pSE211 TraJ which has one predicted transmembrane helix, and a FtsK/SpoIIIE domain at the C-terminal part, except for TraJ of pSE222 which has a RecA-NTPase domain at that position. This domain is present in bacterial conjugation proteins and related DNA transfer proteins involved in type II and type IV secretion (Marchler-Bauer et al., 2007). The C-terminal half of TraJ of pMEA100 also shows similarity (27%) to that of the main transfer proteins of plasmids pSG5 from *Streptomyces ghanaensis* and pSVH1 of *Streptomyces venezuelae*, which also have two putative transmembrane helices and a FtsK/SpoIIIE domain at a similar location. These proteins are septal DNA translocators able to translocate unprocessed double stranded DNA molecules to recipient strains (Reuther et al., 2006). The lack of transfer protein homologues on pSE101 is in agreement with the observation that it does not show conjugal activity (Brown et al., 1988b). However, there is a possibility that the element is mobilisable, but that the proper conditions to observe transfer of pSE101 have not been met. Transfer genes related to those found on the other AICES are also absent on the closely related pSE102.

3.5. amp121-dmt

amp121 (148 codons) and *amp122* (247 codons) encode putative hypothetical proteins for which no significant homologues were found in databases. No specific domains, transmembrane helices, or other structural features were detected, but the

codon usage of these proteins is similar to that of the other pMEA100-encoded ORFs, suggesting that they may encode functional proteins.

A putative single-stranded binding protein (SSB) is encoded downstream of *amp122*. Amp123 is 50% identical to a putative SSB (SACE_1125) encoded on pSE222 (Table 1). SSBs bind a-specifically to ssDNA and have an important role in DNA replication, recombination and repair in bacteria and bacteriophages (Chase and Williams, 1986; Meyer and Laine, 1990). They are present on many conjugative plasmids of several Incompatibility (Inc) groups (Golub and Low, 1985), and presumably prevent a shortage of SSB proteins in recipient strains during transfer of ssDNA intermediates (Jones et al., 1992). Apparently, this is not the function of the pMEA100-encoded SSB, since conjugative transfer of pMEA100 to recipient strains most likely only involves dsDNA, similar to the transfer mechanism of the *Streptomyces* plasmids pSG5 and pSVH1 (Reuther et al., 2006) and pSAM2 (Possoz et al., 2001). The proposed function in pMEA100 is to stabilise and protect large ssDNA intermediates during autonomous replication (te Poele et al., 2006). The other, smaller AICEs do not encode SSB proteins, suggesting that their host SSB pool is sufficient for stabilising the ssDNA intermediates of these elements during replication.

Amp124 and its homologues on pSE211 and pSE222 are discussed in the replication section of this paper.

Amp125 is a 166 aa long conserved hypothetical protein of unknown function with homologues exclusively in the actinomycetes. Amp125 has homologues on pSE211 (SACE_7013), pSE222 (SACE_1136) and SLP1 (SCO4624).

Amp126 (117 aa) is a conserved hypothetical protein of unknown function with homologues on the genome of *S. coelicolor* (SCO4460) and on its linear plasmid SCP1 (SCP1.122) and in a few other unrelated bacterial species.

The final *orf* on pMEA100 encodes a 541 aa protein with high similarity (51% identity) to a putative C5 cytosine-specific DNA methyltransferase (DMT) of *Nocardia farcinica* (pnf2750). C5 cytosine-specific *dmts* are also found on pSE211 and pSE222, and as in pMEA100, the genes are located directly upstream of *repAM*, are related to C5 cytosine-specific DMTs of Mycobacteriophages, and their protein products also contain the highly conserved N-terminal motifs involved in cofactor binding and catalysis of methyl transfer (Kumar et al., 1994). However, the proteins are just 224 and 248 aa long, respectively, and show only weak similarity (~20% identity) to the pMEA100-encoded DMT. AICE*Sco*5349 encodes a 1200 aa putative DMT (SCO5331) with a N6 adenine-specific DNA methylase signature but is not related to the DMTs of the other AICEs. DMTs play an important role in passing the host restriction barrier by methylating a specific base in the target sequences and are found in many bacteriophages and self-transmissible Inc plasmids (Kruger and Bickle, 1983; Posfai et al., 1989).

3.5.1. Element-specific genes

In addition to the commonly shared genes, some genes are present only on one or a few (pMEA-) elements. Element-specific genes are as expected mainly found on the three larger elements pMEA100, pSE222 and pSE211 and mostly encode hypothetical proteins with unknown function.

The AICEs pSAM2, AICE*Sco*5349, and pSE211 encode proteins, ORF183 (183 aa), SCO5345 (183 aa) and SACE_7027 (202 aa), respectively, that are related to each other, but have no homologues on the other elements. ORF183 and SCO5345 show 81% identity and are both ~27% identical to SACE_7027. These putative proteins contain a GGDEF domain that is known to be involved in the synthesis of c-di-GMP. This molecule is emerging as an important second messenger in a recently discovered signalling system that appears to be ubiquitous in bacteria (Romling et al., 2005).

pSE222 encodes a putative flavoprotein of 157 aa (SACE_1142), but the significance of its presence on pSE222 is not clear. pSE222 also contains a gene (SACE_1142a) with a considerable lower G+C content of 51% than the average 66.3%, and its protein product (349 aa) shows 27% (85/310) and 30% (58/189) identity to conserved hypothetical proteins (ZP_01980106; ZP_01705385) of the *Vibrio cholerae* MZO-2 and *Shewanella putrefaciens* 200.

SLP1 encodes a protein of 560 aa (SCO_4631) with a HNH-endonuclease motif. AICE*Sco*5349 also encodes a protein with an HNH-motif (SCO3723 (301 aa), but shows no significant similarity to the SLP1-encoded protein. Proteins with a HNH-signature include transposases, bacterial colicins, restriction endonucleases and homing endonucleases. The latter proteins are encoded by ORFs embedded in introns and inteins, which they mobilise together with themselves by generating strand breaks in cognate sequences (Stoddard,

2005). The function of the AICE-encoded HNH-motif containing proteins is not known.

pMR2 encodes a 243 aa protein GltA with homology to the family of glycosyltransferases involved in cell wall biogenesis.

Interestingly, pSE102 encodes an aminoglycoside phosphotransferase protein (APH). Proteins belonging to this family of phosphotransferase enzymes consist of bacterial antibiotic resistance proteins, which confer resistance to various aminoglycosides, such as streptomycin phosphotransferase (*aphE*) of *S. griseus* that confers resistance to streptomycin (Trower and Clark, 1990). This would be a clear example of a predictable beneficial phenotype to the host, i.e. antibiotic resistance. So far, all described genes on the AICEs appear to be involved in regulation, maintenance and transfer of the element itself. The putative glycosyltransferase of pMR2 and the HNH-protein of SLP1 could be other examples of AICE-encoded genes with a beneficial phenotype to the host. The only observed phenotype that all the AICEs have in common is the pock phenotype upon transfer of the elements. Additionally, mobilisation of chromosomal markers has been observed for several elements, such as pMEA100, pMEA300, pSE211, SLP1, and pSAM2 (Moretti et al., 1985; Vrijbloed, 1996; Brown et al., 1988b; Bibb et al., 1981; Smokvina et al., 1988).

3.6. Comparison of the pMEA-elements

Comparison of the nucleotide and amino acid sequences of the pMEA-related elements reveals several highly conserved regions that may indicate a common origin. In particular, the *repAM* genes of pMEA300 and pSE211 are highly similar, showing 70% nucleotide identity and 61% aa identity. The *xis* gene and protein of pMEA100 show 59% nucleotide identity and 62% aa identity, respectively to *xis* and its protein of pSE211. Also the integrases of pMEA100, pSE211 and pSE101 are highly conserved showing 63 to 65% identity on DNA level and 58 to 63% identity on protein level to each other. pMEA100 and pSE211 share a highly similar DNA region (62%), consisting of a *mdp*, *nud*, *xre* gene, and is not found on the other two elements. Also pMEA300 and pSE101 share a DNA region with high sequence similarity (67%), containing a putative *stf* gene (70% aa identity between both Stf proteins) and a Nudix hydrolase gene (64% aa identity between the two enzymes). The pMEA300 and pSE101 Nudix genes are not related to those of pMEA100 and pSE211.

Alignment of the DNA sequence of pSE101 and pSE102 shows that a large part of pSE102 is almost identical (86% nucleotide identity) to the region of pSE101. This part of the element is probably the result of a recent duplication of pSE101. This region includes putative replication and excisionase genes and the pMEA-unique hairpin. An integrase, *korA*, and the small pMEA-specific ORF (SACE_0248) can be found at a similar position on the element as on pSE101, but they show no significant nucleotide identity to the corresponding genes of pSE101. The elements are also integrated in different tRNA genes; pSE101 in a Thr-tRNA gene and pSE102 in a Ser-tRNA gene. Moreover, pSE102 lacks the DNA region containing SACE_0474, *nud101* and SACE_0476 found on pSE101 and has SACE_0249 and an *aph* gene at that location instead.

It is unknown whether pSE102, like pSE101 (Brown et al., 1988a), can be present in a free replicating form. Since pSE102 encodes a putative replication protein and excisionase that are highly related (92 and 90% identity, respectively) to those of pSE101, and in view of the presence of a pMEA-like integrase and attachment site, we suggest that pSE102 can replicate autonomously.

Comparison of the G+C content of the genes of the pMEA-elements (data not shown) shows that especially genes involved in regulation have a consistently lower G+C content, i.e. genes, *mdp*, *nud1, amp109* and *amp110* of pMEA100, and the corresponding genes on the other elements. Also the integrase genes, except that of pSE222, show a consistently lower G+C content than the average value. Possibly, these genes were acquired recently by horizontal gene transfer from an organism with lower G+C content.

4. Conclusions

AICEs are present in diverse genera of the actinomycetes, encompassing the most important bacterial producers of bioactive secondary metabolites, including many medically and commercially important antibiotics (Bérdy, 2005). This study reveals the presence of two novel putative AICEs in the *S. erythraea* genome in addition to the previously described pSE211 and pSE101. The two elements, named pSE222 and pSE102, have features in common with the pMEA-elements and the other AICEs. Homologues of several AICE-encoded proteins

were found in other actinomycetes, like in *Frankia* and the obligate marine genus *Salinispora*. Preliminary results show that they might be part of AICEs as well.

AICE*Sco*5349 and AICE*Sco*3250 are two of several potentially horizontal acquired insertions in the *S. coelicolor* genome (Bentley et al., 2002). Interestingly, AICE*Sco*5349 flanks the *whiE* cluster (Bentley et al., 2002) encoding the enzymes to synthesise the grey polyketide spore pigment (Davis and Chater, 1990) and AICE*Sco*3250 flanks the calcium-dependent antibiotic (CDA) biosynthetic gene cluster. The two AICEs are absent on the highly syntenic core region of the *S. lividans* genome (Jayapal et al., 2007) and the *S. ambofaciens* genome lacks the CDA cluster and also the complete region containing the *whiE* cluster and AICE*Sco*5349 (Choulet et al., 2006). Possibly, these AICEs are associated with the acquisition of these secondary metabolite clusters in *S. coelicolor*.

Several proteins of the *Amycolatopsis* and *Saccharopolyspora* AICEs are also related to proteins encoded by other potentially recently laterally acquired insertions in the *S. coelicolor* genome. For instance, the integrase of pSE222 is 69% identical to the integrase of a large insertion that includes genes for heavy metal resistance and secondary metabolite production. Conceivably, these integrases have a common origin.

Interestingly, a putative aminoglycoside phosphotransferase gene (*aph*) was found on pSE102. Such genes are often present in aminoglycoside antibiotic biosynthesis clusters, acting as resistance genes, protecting the host against the negative effects of the antibiotic it produces. Genes encoding clear beneficial functions for the physiology of the hosts are not, or very rarely, found on these elements. The aminoglycoside phosphotransferase gene found on pSE102 could be a clear example of a predictable beneficial phenotype to the host, i.e. antibiotic resistance.

Although not shown thus far, it has been suggested that the genes responsible for the widespread resistance to aminoglycoside antibiotics in pathogens originate from soil bacteria, notably the actinomycetes, which are the most prominent aminoglycoside producers, and that the resistance genes were spread to other bacteria via horizontal gene transfer (Benveniste and Davies, 1973; Marshall et al., 1998; Grohmann et al., 2003). Several AICEs are able to mobilise chromosomal markers (Moretti et al., 1985; Hopwood et al., 1984; Brown et al., 1988b; Vrijbloed, 1996; Smokvina et al., 1988; Bibb et al., 1981), suggesting that they may play an important role in horizontal gene transfer and evolution of genome plasticity. In view of the above-mentioned observations, it is tempting to speculate that these AICEs are involved in spread of antibiotic resistance as well.

Acknowledgments

We thank Jolanda Brons for DNA nucleotide sequencing. P.L. thanks the Biotechnology and Biological Sciences Research Council (UK) for support of this work (in part).

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Aravind, L., Koonin, E.V., 1998. The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends Biochem. Sci. 23, 469–472.

Babcock, M.J., Kendrick, K.E., 1990. Transcriptional and translational features of a sporulation gene of *Streptomyces griseus*. Gene 95, 57–63.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., et al., 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417, 141–147.

Benveniste, R., Davies, J., 1973. Aminoglycoside antibiotic-inactivating enzymes in actinomycetes similar to those present in clinical isolates of antibiotic-resistant bacteria. Proc. Natl. Acad. Sci. USA 70, 2276–2280.

Bérdy, J., 2005. Bioactive microbial metabolites. J. Antibiot. (Tokyo) 58, 1–26.

Bibb, M.J., Ward, J.M., Kieser, T., Cohen, S.N., Hopwood, D.A., 1981. Excision of chromosomal DNA sequences from *Streptomyces coelicolor* forms a novel family of plasmids detectable in *Streptomyces lividans*. Mol. Gen. Genet. 184, 230–240.

Boccard, F., Smokvina, T., Pernodet, J.L., Friedmann, A., Guerineau, M., 1989. The integrated conjugative plasmid pSAM2 of *Streptomyces ambofaciens* is related to temperate bacteriophages. EMBO J. 8, 973–980.

Brasch, M.A., Cohen, S.N., 1993. Excisive recombination of the SLP1 element in *Streptomyces lividans* is mediated by Int and enhanced by Xis. J. Bacteriol. 175, 3075–3082.

Brown, D.P., Chiang, S.J., Tuan, J.S., Katz, L., 1988a. Site-specific integration in *Saccharopolyspora erythraea* and multisite integration in *Streptomyces lividans* of actinomycete plasmid pSE101. J. Bacteriol. 170, 2287–2295.

Brown, D.P., Tuan, J.S., Boris, K.A., DeWitt, J.P., Idler, K.B., Chiang, S.J., Katz, L., 1988b. Plasmid-chromosome interactions in *Saccharopolyspora erythraea* and *Streptomyces lividans*. Dev. Ind. Microbiol. 29, 97–105.

Brown, D.P., Idler, K.B., Backer, D.M., Donadio, S., Katz, L., 1994. Characterization of the genes and attachment sites for site-specific integration of plasmid pSE101 in *Saccharopolys-*

pora erythraea and *Streptomyces lividans*. Mol. Gen. Genet. 242, 185–193.

Brown, D.P., Idler, K.B., Katz, L., 1990. Characterization of the genetic elements required for site-specific integration of plasmid pSE211 in *Saccharopolyspora erythraea*. J. Bacteriol. 172, 1877–1888.

Chase, J.W., Williams, K.R., 1986. Single-stranded DNA binding proteins required for DNA replication. Annu. Rev. Biochem. 55, 103–136.

Choulet, F., Aigle, B., Gallois, A., et al., 2006. Evolution of the terminal regions of the *Streptomyces* linear chromosome. Mol. Biol. Evol. 23, 2361–2369.

Davis, N.K., Chater, K.F., 1990. Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. Mol. Microbiol. 4, 1679–1691.

Frost, L.S., Leplae, R., Summers, A.O., Toussaint, A., 2005. Mobile genetic elements: the agents of open source evolution. Nat. Rev. Microbiol. 3, 722–732.

Golub, E.I., Low, K.B., 1985. Conjugative plasmids of enteric bacteria from many different incompatibility groups have similar genes for single-stranded DNA-binding proteins. J. Bacteriol. 162, 235–241.

Grohmann, E., Muth, G., Espinosa, M., 2003. Conjugative plasmid transfer in gram-positive bacteria. Microbiol. Mol. Biol. Rev. 67, 277–301.

Hagege, J., Pernodet, J.L., Sezonov, G., Gerbaud, C., Friedmann, A., Guerineau, M., 1993. Transfer functions of the conjugative integrating element pSAM2 from *Streptomyces ambofaciens*: characterization of a *kil-kor* system associated with transfer. J. Bacteriol. 175, 5529–5538.

Hopwood, D.A., Hintermann, G., Kieser, T., Wright, H.M., 1984. Integrated DNA sequences in three streptomycetes form related autonomous plasmids after transfer to *Streptomyces lividans*. Plasmid 11, 1–16.

Iyer, L.M., Koonin, E.V., Leipe, D.D., Aravind, L., 2005. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. Nucleic Acids Res. 33, 3875–3896.

Jayapal, K.P., Lian, W., Glod, F., Sherman, D.H., Hu, W.S., 2007. Comparative genomic hybridizations reveal absence of large *Streptomyces coelicolor* genomic islands in *Streptomyces lividans*. BMC Genomics 8, 229.

Jones, A.L., Barth, P.T., Wilkins, B.M., 1992. Zygotic induction of plasmid ssb and psiB genes following conjugative transfer of Incl1 plasmid Collb-P9. Mol. Microbiol. 6, 605–613.

Kruger, D.H., Bickle, T.A., 1983. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. Microbiol. Rev. 47, 345–360.

Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J., Wilson, G.G., 1994. The DNA (cytosine-5) methyltransferases. Nucleic Acids Res. 22, 1–10.

Lipps, G., 2004. The replication protein of the *Sulfolobus islandicus* plasmid pRN1. Biochem. Soc. Trans. 32, 240–244.

Lipps, G., Rother, S., Hart, C., Krauss, G., 2003. A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. EMBO J. 22, 2516–2525.

Madon, J., Moretti, P., Hutter, R., 1987. Site-specific integration and excision of pMEA100 in *Nocardia mediterranei*. Mol. Gen. Genet. 209, 257–264.

Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., et al., 2007. CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res. 35, D237–D240.

Marshall, C.G., Lessard, I.A., Park, I., Wright, G.D., 1998. Glycopeptide antibiotic resistance genes in glycopeptide-producing organisms. Antimicrob. Agents Chemother. 42, 2215–2220.

McGuire, J.M., Bunch, R.L., Anderson, R.C., Boaz, H.E., Flynn, E.H., Powell, H.M., Smith, J.W., 1952. Ilotycin, a new antibiotic. Antibiot. Chemother. 2, 281–284.

McLennan, A.G., 2006. The Nudix hydrolase superfamily. Cell Mol. Life Sci. 63, 123–143.

Meyer, R.R., Laine, P.S., 1990. The single-stranded DNA-binding protein of *Escherichia coli*. Microbiol. Rev. 54, 342–380.

Moretti, P., Hintermann, G., Hutter, R., 1985. Isolation and characterization of an extrachromosomal element from *Nocardia mediterranei*. Plasmid 14, 126–133.

Oliynyk, M., Samborskyy, M., Lester, J.B., Mironenko, T., Scott, N., Dickens, S., Haydock, S.F., Leadlay, P.F., 2007. Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. Nat. Biotechnol. 25, 447–453.

Posfai, J., Bhagwat, A.S., Posfai, G., Roberts, R.J., 1989. Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res. 17, 2421–2435.

Possoz, C., Gagnat, J., Sezonov, G., Guerineau, M., Pernodet, J.L., 2003. Conjugal immunity of *Streptomyces* strains carrying the integrative element pSAM2 is due to the pif gene (pSAM2 immunity factor). Mol. Microbiol. 47, 1385–1393.

Possoz, C., Ribard, C., Gagnat, J., Pernodet, J.L., Guerineau, M., 2001. The integrative element pSAM2 from *Streptomyces*: kinetics and mode of conjugal transfer. Mol. Microbiol. 42, 159–166.

Reuther, J., Gekeler, C., Tiffert, Y., Wohlleben, W., Muth, G., 2006. Unique conjugation mechanism in mycelial streptomycetes: a DNA-binding ATPase translocates unprocessed plasmid DNA at the hyphal tip. Mol. Microbiol. 61, 436–446.

Rigali, S., Derouaux, A., Giannotta, F., Dusart, J., 2002. Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. J. Biol. Chem. 277, 12507–12515.

Romling, U., Gomelsky, M., Galperin, M.Y., 2005. C-di-GMP: the dawning of a novel bacterial signalling system. Mol. Microbiol. 57, 629–639.

Sensi, P., Greco, A.M., Ballotta, R., 1959. Rifomycin I. Isolation and properties of rifomycin B and rifomycin complex. Antibiot. Annu. 7, 262–270.

Sezonov, G., Possoz, C., Friedmann, A., Pernodet, J.L., Guerineau, M., 2000. KorSA from the *Streptomyces* integrative element pSAM2 is a central transcriptional repressor: target genes and binding sites. J. Bacteriol. 182, 1243–1250.

Shiffman, D., Cohen, S.N., 1993. Role of the *imp* operon of the *Streptomyces coelicolor* genetic element SLP1: two imp-encoded proteins interact to autoregulate imp expression and control plasmid maintenance. J. Bacteriol. 175, 6767–6774.

Smokvina, T., Boccard, F., Pernodet, J-L., Friedmann, A., Guerineau, M., 1991. Functional analysis of the *Streptomyces ambofaciens* element pSAM2. Plasmid 25, 40–52.

Smokvina, T., Francou, F., Luzzati, M., 1988. Genetic analysis in *Streptomyces ambofaciens*. J. Gen. Microbiol. 134, 395–402.

Stoddard, B.L., 2005. Homing endonuclease structure and function. Q. Rev. Biophys. 38, 49–95.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol. Biol. Evol..

Tan, G.Y.A., Ward, A.C., Goodfellow, M., 2006. Exploration of *Amycolatopsis* diversity in soil using genus-specific primers and novel selective media. Syst. Appl. Microbiol. 29, 557–569.

te Poele, E.M., Habets, M.N., Tan, G.Y., Ward, A.C., Goodfellow, M., Bolhuis, H., Dijkhuizen, L., 2007. Prevalence and distribution of nucleotide sequences typical for pMEA-like accessory genetic elements in the genus *Amycolatopsis*. FEMS Microbiol. Ecol. 61, 285–294.

te Poele, E.M., Kloosterman, H., Hessels, G.I., Bolhuis, H., Dijkhuizen, L., 2006. RepAM of the *Amycolatopsis methanolica* integrative element pMEA300 belongs to a novel class of replication initiator proteins. Microbiology 152, 2943–2950.

Toussaint, A., Merlin, C., 2002. Mobile elements as a combination of functional modules. Plasmid 47, 26–35.

Trower, M.K., Clark, K.G., 1990. PCR cloning of a streptomycin phosphotransferase (aphE) gene from *Streptomyces griseus* ATCC 12475. Nucleic Acids Res. 18, 4615.

van Wezel, G.P., van der, M.J., Kawamoto, S., Luiten, R.G., Koerten, H.K., Kraal, B., 2000. ssgA is essential for sporulation of *Streptomyces coelicolor* A3(2) and affects hyphal development by stimulating septum formation. J. Bacteriol. 182, 5653–5662.

Vrijbloed, J.W., 1996. Functional analysis of the integrative plasmid pMEA300 of the actinomycete *Amycolatopsis methanolica*. PhD Thesis, University of Groningen.

Vrijbloed, J.W., Jelinkova, M., Hessels, G.I., Dijkhuizen, L., 1995a. Identification of the minimal replicon of plasmid pMEA300 of the methylotrophic actinomycete *Amycolatopsis methanolica*. Mol. Microbiol. 18, 21–31.

Vrijbloed, J.W., Madon, J., Dijkhuizen, L., 1994. A plasmid from the methylotrophic actinomycete *Amycolatopsis methanolica* capable of site-specific integration. J. Bacteriol. 176, 7087–7090.

Vrijbloed, J.W., Madon, J., Dijkhuizen, L., 1995b. Transformation of the methylotrophic actinomycete *Amycolatopsis methanolica* with plasmid DNA: stimulatory effect of a pMEA300-encoded gene. Plasmid 34, 96–104.

Vrijbloed, J.W., van der Put, N.M., Dijkhuizen, L., 1995c. Identification and functional analysis of the transfer region of plasmid pMEA300 of the methylotrophic actinomycete *Amycolatopsis methanolica*. J. Bacteriol. 177, 6499–6505.

Wink, J.M., Kroppenstedt, R.M., Ganguli, B.N., Nadkarni, S.R., Schumann, P., Seibert, G., Stackebrandt, E., 2003. Three new antibiotic producing species of the genus *Amycolatopsis*, *Amycolatopsis balhimycina* sp. nov., *A. tolypomycina* sp. nov., *A. vancoresmycina* sp. nov., and description of *Amycolatopsis keratiniphila* subsp. *keratiniphila* subsp. nov. and *A. keratiniphila* subsp. *nogabecina* subsp. nov. Syst. Appl. Microbiol. 26, 38–46.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.